Understanding Parental Characteristics of Child Adoption Candidates using MMPI-2 and Evolutionary Clustering

Emanuele Iacobelli[1], Cristian Randieri[2], Paolo Roma[3] and Samuele Russo[4]

[1]Department of Computer, Control and Management Engineering, Sapienza University of Rome, 00185 Roma, Italy;
[2]Università degli Studi eCampus, Novedrate (CO), Italy;
[3]Department of Human Neuroscience, Sapienza University of Rome, Italy;
[4]Department of Psychology, Sapienza University of Rome, Italy;

Abstract

In the context of adoption, evaluating prospective adoptive parents using psychometric assessments such as the Minnesota Multiphasic Personality Inventory (MMPI) questionnaire is essential for understanding their psychological profiles. However, interpreting such complex data can be both challenging and time-consuming. In this study, we propose a meta-analysis tool to assist psychologists in their initial interpretation and analysis of MMPI-2 results by providing a clear data-driven visualization of key psychometric scales. Our system employs unsupervised learning techniques to uncover meaningful patterns and relationships in the data with minimal prior input. Specifically, a genetic algorithm is used to optimize clustering quality by selecting the most relevant psychological scales, enhancing cluster separation, and improving data interpretability. We also explored and compared the effectiveness of several clustering algorithms, including K-Means, Gaussian Mixture Model, and Spectral Clustering, to maximize the capabilities of our tool.

Keywords

Minnesota Multiphasic Personality Inventory (MMPI), Unsupervised Learning Algorithms, Genetic Algorithm, K-mean, Gaussian Mixture Model, Spectral Clustering

1. Introduction

Adoption is the process whereby individuals or families assume the parenting of a child who is not biologically their own. According to specific studies [1, 2, 3, 4], sometimes adoptees could have problems in psychological development, social relationships, and establishing a sense of identity. Therefore, finding suitable adoptive parents is crucial for the well-being of the child.

For that reason, standardized psychometric tests [5, 6, 7, 8] are used to assess the personality and psychopathology traits of prospective adoptive parents. An example of such a test is the Minnesota Multiphasic Personality Inventory (MMPI) psychological test [9], proposed in 1943. Over the years, several variations of the test have been developed. The most commonly used versions today include the MMPI-2 [10], which was published in 1989 specifically for adults; the MMPI-A [11], designed for adolescents and introduced in 1992; the MMPI-Restructured Form, a condensed version of the MMPI; and the recently released MMPI-3 [12], published in 2020.

For the evaluation of the results, the set of most important psychometric scales to be analyzed is usually handpicked by field experts as it is highly task-dependent. For that reason, in this study, we propose an unsupervised learning algorithm capable of clustering the data gathered with the MMPI-2 test using as little as possible prior knowledge during the preprocessing and postprocessing of the data.

The clustering [13] process is an unsupervised learning technique designed to identify similarities within data without predefined categories. In our case, by analyzing the geometric properties of the data, the goal is to capture as many similarities as possible, even when the underlying distribution is not known a priori. Our approach involves the development of a machine learning based [14, 15, 16, 17] genetic algorithm [18, 19, 20, 21] aimed at optimizing both the minimum centroid distance and the minimum inter-cluster distance, enhancing the clustering quality. We also conducted experiments with three different clustering algorithms (K-Means [22, 23, 24, 25], Gaussian mixture model [26, 27], and Spectral clustering [28]) to determine the most suitable one for our system. In particular, given that the number of clusters is not predetermined, careful interpretation of the results is necessary to attribute meaningful explanations to each cluster.

ICYRIME 2024: 9th International Conference of Yearly Reports on Informatics, Mathematics, and Engineering. Catania, July 29-August 1, 2024

iacobelli@diag.uniroma1.it (E. Iacobelli);
cristian.randieri@uniecampus.it (C. Randieri);
paolo.roma@uniroma1.it (P. Roma); samuele.russo@uniroma1.it (S. Russo)
0009-0003-1379-9106 (E. Iacobelli); 0000-0001-5300-3561 (C. Randieri); 0000-0002-1031-0948 (P. Roma); 0000-0002-1846-9996 (S. Russo)

© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR-WS.org/Vol-3869/p09.pdf

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

1.1. Roadmap

This paper is organized as follows: first, an overview of the MMPI-2 questionnaire, its scales, and traditional MMPI clustering methods is presented in Section 2. Next, Section 3 provides a detailed description of the core techniques used in our algorithm. In Section 4, we describe the dataset employed in this experiment. Following this, Section 5 offers a comprehensive explanation of the system we developed and its evaluation process. The clustering results produced by our system are then presented in Section 6. Finally, Section 7 summarizes the article's content and outlines potential areas for future improvement.

2. State of the Art

2.1. MMPI-2 Overview

The MMPI-2 is used as a personality assessment tool in clinical and non-clinical contexts to discern psychopathologies and behavioral traits in individuals. It comprises a series of true/false questions, known as items, which are grouped into various scales designed to measure specific aspects of the subject's disposition.

Validity scales scrutinize the subject's approach to the test and demeanor, identifying inconsistencies or attempts to manipulate responses. Among them, the Lie scale (L) evaluates honesty during the test, while the K scale assesses defensive tendencies and reluctance to acknowledge personal issues.

In addition, the MMPI incorporates ten primary clinical scales designed to detect a spectrum of psychological disorders, encompassing Hypochondriasis (Hs), Depression (D), Hysteria (Hy), Psychopathic Deviate (Pd), Masculinity/Femininity (Mf), Paranoia (Pa), Psychasthenia (Pt), Schizophrenia (Sc), Hypomania (Ma), and Social Introversion (Si). Furthermore, content scales target specific personal attitudes, including anger issues (ANG), low self-esteem (LSE), family problems (FAM), and work-related challenges (WRK), among others.

Additionally, supplemental scales are used in combination with the content scales to determine if some symptoms are attributed to alternative potential causes such as controlled hostility, alcoholism, and more.

Moreover, Psy-5 scales measure dimensional traits of personality disorders, including Aggressiveness, Psychoticism, Constraint, Neuroticism, and Extraversion.

Finally, to ensure uniform interpretation across all scales, scores are transformed into T-scores, ranging from 30 to 120. Typically, scores exceeding 65 are considered significant and warrant further examination.

2.2. Traditional MMPI Clustering Methods

Following this concise overview of the MMPI-2 test, prior attempts to cluster datasets derived from this assessment have typically involved manually selecting sets of the aforementioned psychometric scales.

In [29], an algorithm very similar to K-Means (originally described in [30]) was applied to data obtained from MMPI-2 tests administered to women in their third trimester of pregnancy. The objective was to determine the personality characteristics of women who develop perinatal depression.

Similarly, in [31], clusters were generated to identify groups of chronic low-back pain patients based on personality traits identified through the MMPI-2 test.

Another notable study is presented in [32], where the authors investigated individuals trained to simulate Post-traumatic Stress Disorder (PTSD). They conducted cluster analysis on MMPI-2 clinical and validity scales, identifying two well-fitting cluster solutions. Discriminant and multivariate analyses of variance (MANOVAs) were employed to evaluate the clusters, revealing significant differences in MMPI-2 content scales. Specifically, demographic variables had minimal influence on cluster membership, but there were discrepancies in the reported clarity of PTSD education materials among clusters.

In [33], the authors investigated the MMPI-2-RF validity scales' effectiveness in profiling chronic pain patients. To identify clusters, a two-step exploratory cluster analysis was conducted, employing the auto-clustering selection feature in IBM SPSS 21 to select the optimal cluster solution. Cluster analysis revealed two distinct patient clusters. Cluster 1 displayed valid responses and exhibited elevations primarily on somatic and low positive emotion scales. In contrast, Cluster 2 comprised patients who overreported on validity scales and demonstrated elevations on multiple restructured clinical scales.

3. Core Techniques in Our Algorithm

3.1. Genetic Algorithm

All cited works in this paper employ clustering techniques with input from psychology experts to select relevant psychometric scales for analysis. In contrast, our system autonomously selects key scales using a genetic algorithm [34]. Genetic algorithms (GAs) are adaptive search procedures widely utilized in Artificial Intelligence since the 1970s [35, 36, 37]. Drawing inspiration from biological evolution, GAs simulate aspects of the process of natural selection proposed by Charles Darwin. They involve successive generations of candidate solu-

tions undergoing reproduction, mutation, and selection to converge toward optimal or near-optimal solutions. Genetic algorithms have a broad range of applications [38, 39, 40]; any problem that can be formalized as a string of 0s and 1s can potentially be optimized using this approach.

In summary, a general genetic algorithm workflow is the following: firstly, an initial population of individuals (each represented as a string of 0s and 1s) is randomly generated. Next, a fitness value is assigned to each individual in the population according to a certain fitness function. Then, multiple pools of individuals are randomly selected, and a certain number of individuals are chosen based on their fitness value to serve as parents for the next population from each pool. For each pair of parents, two children are produced using the following criteria: a crossover index is randomly selected and determines how much of the first part of one parent's string is merged with the second part of the other parent's string, and vice versa. Finally, each bit of the generated children is flipped according to a certain probability simulating the mutation process. This algorithm continues until a specific number of consecutive iterations occur without any improvement in the best fitness value. When the algorithm halts, the latest best individual found is selected as the optimal solution discovered thus far.

3.2. Clustering Algorithms

Clustering algorithms belong to the unsupervised learning domain of artificial intelligence and are designed to unveil concealed patterns and organize data points into coherent clusters based on their intrinsic similarities. These algorithms rely on different distance metrics like Euclidean distance, cosine similarity, and the Jaccard coefficient to quantify the resemblance between data points. The typical representation of each resulting cluster involves a centroid, acting as a central reference point summarizing the collective traits of its constituent data points. These algorithms can be broadly categorized into several methodologies. Partitioning methods, exemplified by K-means, iteratively segment the dataset into non-overlapping clusters, ensuring each data point exclusively belongs to one cluster. Hierarchical methods, such as Agglomerative clustering, construct a hierarchical arrangement of clusters by iteratively merging or dividing existing clusters based on similarity criteria, culminating in a tree-like structure. Model-based methods, on the other hand, assume that the data is generated by a probabilistic model, such as a Gaussian Mixture Model (GMM), allowing for the probabilistic modeling of clusters.

In our study, we focus on evaluating and comparing the performance of K-means, Gaussian Mixture Model, and Spectral Clustering.

In detail, K-Means partitions samples into a predefined number of clusters through an iterative process: randomly selecting K samples as initial clusters (and centroids), assigning each sample to the cluster with the nearest centroid, recomputing centroids, and terminating the process if no data points have switched clusters or if the distance between new and old centroids falls below a certain threshold.

Gaussian Mixture Model (GMM) endeavors to fit a specified number (N) of normal distributions to distinct subsets of the original dataset by estimating their mean and variance parameters using the Expectation-Maximization (EM) algorithm [41].

Spectral Clustering, on the other hand, exploits the spectral properties of the affinity matrix to capture the underlying data structure, particularly in scenarios where traditional clustering techniques may struggle with nonlinear or intricate relationships between data points. In particular, it leverages techniques such as spectral decomposition (eigenvalue decomposition) or singular value decomposition (SVD), to transform data into a lower-dimensional space and subsequently employs a standard clustering algorithm, such as K-means, to partition the data points into clusters.

4. Dataset

In this study, we utilized a dataset comprising 202 entries and 813 features for each entry. These features encompass anamnestic information, boolean answers to the MMPI's questions, and T-scores. Figures 1, 2, and 3 provide an overview of the statistics regarding some of the anamnestic information and the clinical and content scales, calculated as T-scores, of the subjects in our dataset. For preprocessing, we removed features with either a single value or a predominant value (e.g., 'Citizenship') and those with high variability (e.g., 'Profession'). Additionally, we dropped the gender column since MMPI scales have the same interpretation for both men and women. The boolean answers to the MMPI's questions were also discarded, as the normalized T-score values automatically encode this information.

To ensure data validity, according to the guidelines provided by the authors of the MMPI test, applicants with Lie scale scores exceeding 75 were excluded. Additionally, none of the test-takers reached the cutoff of 30 unanswered questions on the 'cannot say' scale that should invalidate the test. We also examined other validity scales such as F, TRINT, and VRINT, but no entries were excluded based on these scales. Applicants with high values indicating alcohol or drug issues were marked as rejected in advance.

The remaining data, consisting of 191 entries with 120 feature columns, was scaled to ensure all features had the same magnitude within the range [0,1]. This scaling

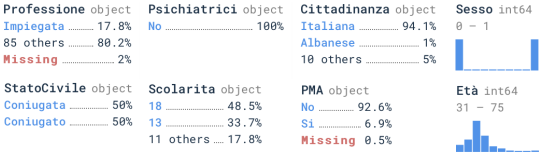

Figure 1: This image displays some of the anamnestic information found in our dataset. From the top left to the bottom right, we have: 'Profession', 'Psychiatric Patients', 'Citizenship', 'Gender', 'Marital Status', 'Education', 'PMA', and 'Age'.

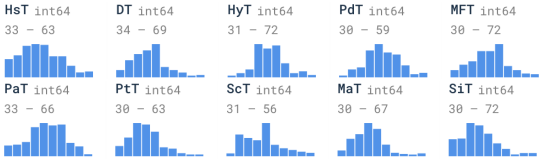

Figure 2: This image displays statistics for various clinical scales, calculated as T-values, found in the dataset.

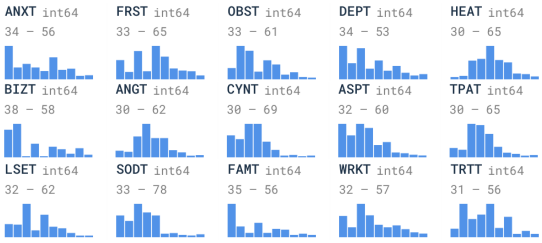

Figure 3: This image displays statistics for various content scales, calculated as T-values, found in the dataset.

was crucial to prevent the overwhelming importance of certain features, particularly the MMPI scales, compared to the boolean values.

5. Methodology and System's Evaluation

For clustering the dataset using a genetic algorithm, each feature in our dataset has been encoded with a binary digit [0,1]. This encoding allows each individual to represent a unique combination of features. Features assigned the value 1 will be considered in the clustering process, while those denoted with 0 will be discarded. Each individual is then evaluated using two different fitness functions: the minimum inter-cluster distance and the minimum centroid distance.

The minimum inter-cluster distance calculates the minimum distance between two data points belonging to different clusters through the following formula:

Table 1
This table presents the results achieved by combining various clustering algorithms and fitness functions within our genetic algorithm applied to a synthetically generated dataset (an example is displayed in Fig. 4). The 'Accuracy' column represents the proportion of correctly classified data points, while the 'Iteration' column indicates the number of iterations the algorithm took to achieve the best result.

Clustering Algorithm	Fitness Function	Accuracy	Iteration
K-Means	Minimum Inter-Cluster Distance	79,5%	6
K-Means	Minimum Centroid Distance	81.4%	5
GMM	Minimum Inter-Cluster Distance	68,8%	5
GMM	Minimum Centroid Distance	72,6%	7
Spectral Analysis	Minimum Inter-Cluster Distance	62,3%	23
Spectral Analysis	Minimum Centroid Distance	64,7%	25

$$V = \min_{\substack{s_i \in C_i \\ s_j \in C_i \\ i \neq j}} \{d(s_i, s_j)\} \tag{1}$$

where s_i and s_j are two distinct data points belonging to different clusters C_i and C_j, respectively, and $d(.,.)$ represents the Euclidean distance function.

The minimum centroid distance measures the distance between the centroids of different clusters through the following formula:

$$V = \min_{i \neq j} \{d(c_i, c_j)\}, c_i = \frac{\sum_{s_i \in C_i} s_i}{|C_i|} \tag{2}$$

where s_i and c_i represents a data point and the centroid of the cluster C_i, respectively, and $d(.,.)$ denotes the Euclidean distance function.

To determine the best combination of the clustering algorithm and fitness function, we evaluated all their possible combinations on a synthetically generated dataset. This dataset was generated by sampling data points from three normal distributions with closely located centroids and large variance, making the clustering more challenging. Specifically, we used three 250-dimensional Gaussian distributions with random means in the range [-1.25,1.25] and a standard deviation equal to 20. To visualize the synthetic dataset in two dimensions (refer to Fig. 4 for an example of the data that can be produced), we applied the Principal Component Analysis (PCA) dimensionality reduction algorithm [42]. The best results were achieved by combining K-means with minimum centroid distance, resulting in an accuracy of 81,4%. Results from other combinations are presented in Table 1, while in Fig. 5 a visual representation of the results is proposed.

6. Results

To determine the optimal number of clusters for the K-Means clustering algorithm on the analyzed dataset, we

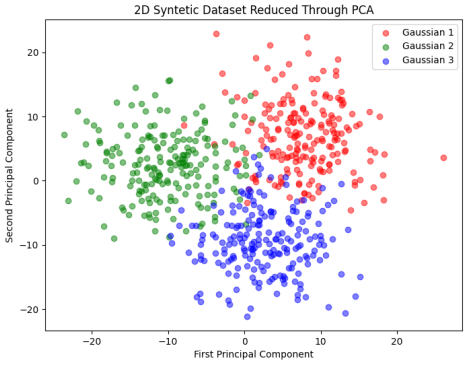

Figure 4: Example of a 2D synthetically generated dataset that we have used to evaluate the best combination of clustering algorithm and fitness function for our genetic algorithm. This dataset was sampled from three 250-dimensional Gaussian distributions with random means in the range [-1.25, 1.25] and a standard deviation of 20. To visualize the data in 2D, we applied the Principal Component Analysis (PCA) algorithm to reduce the dimensionality.

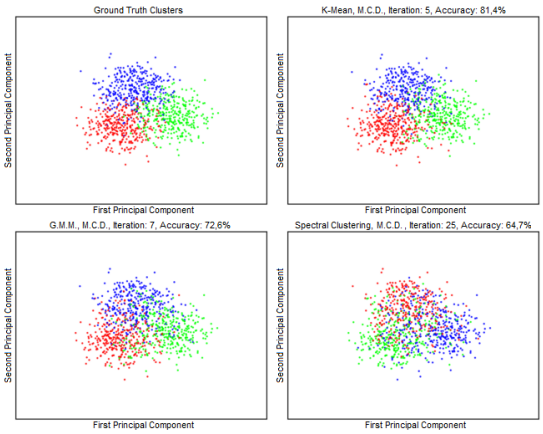

Figure 5: This image represents the accuracy obtained by different clustering algorithms tested on a synthetic dataset to determine the most suitable algorithm for our work. Specifically, we compared K-means, Gaussian Mixture Model, and Spectral Clustering, all using the Minimum Centroid Distance as the fitness function. The results showed that K-means was the best algorithm, achieving an accuracy of 81.4% compared to the ground truth.

employed the Silhouette Analysis. This technique involves computing the Silhouette Coefficient s for each element in the dataset, defined by:

$$s = \frac{b - a}{\mathbf{max}(a, b)} \tag{3}$$

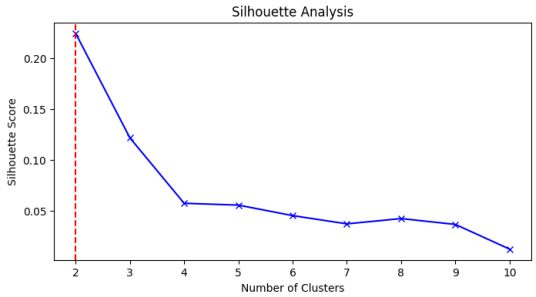

Figure 6: This image represents the Silhouette scores (y-axis) obtained by our algorithm using different numbers of clusters (x-axis). Higher Silhouette scores indicate denser and better-separated clusters. For our dataset, the optimal score was achieved using 2 clusters, as highlighted by the vertical red dashed line.

where a is the mean distance between a sample and all other points in the same cluster, and b is the mean distance between a sample and all points in the nearest cluster. The Silhouette Score, which is the average of the Silhouette Coefficients for all elements in the dataset, indicates the quality of clustering. A higher mean Silhouette Score suggests denser and better-separated clusters. In our study, the optimal number of clusters found for our dataset was 2, as shown in Fig. 6. The Fig. 7 provides a comprehensive overview of Silhouette coefficients for different numbers of clusters, demonstrating the decline in clustering quality as the number of clusters increases.

Executing the PCA to the obtained clusters generates the plot displayed in Fig. 8. It can be seen that on the first principal component (x-axis) the two clusters are well distinguished while on the second principal component (y-axis) they both spread homogeneously even if the elements belonging to the green cluster are more concentrated around the zero value of that axis.

In a more detailed analysis, Fig. 9 illustrates the intra-cluster average values for the four main group scales: Validity, Clinical, Content, and Supplemental. As observed, the elements in the green cluster consistently show lower average values compared to those in the red cluster, with the exception of the Validity scale. This reversal in trend may prompt psychologists to further examine these two clusters, as the scales within the Validity group are designed to indicate how reliable and truthful the test responses are. However, the differences between the clusters are minor, and both demonstrate a high level of reliability in responses, with few outliers. One of the key insights from this analysis is the notable difference in the Content scale, suggesting that individuals in the red cluster may exhibit more psychological issues compared to those in the green cluster.

A similar trend, observed in Fig. 8, is also highlighted

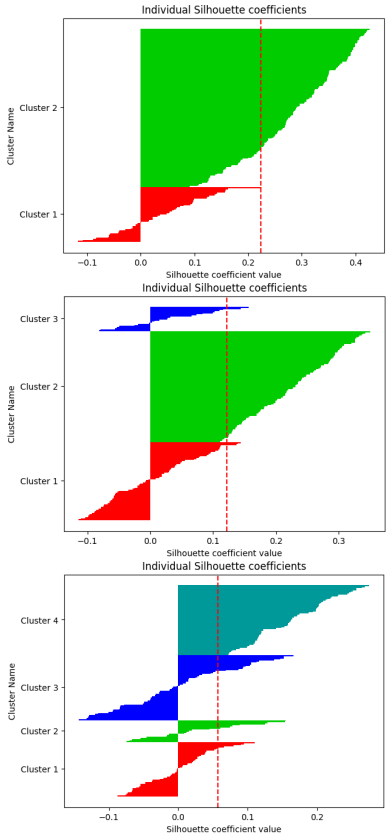

Figure 7: Starting from the top, each plot in this image represents the Silhouette coefficients of all the elements in the dataset, obtained by our algorithm using 2, 3, and 4 clusters, respectively. The y-axis displays the dataset elements divided by the cluster to which they belong, while the x-axis shows the Silhouette coefficient. The vertical red dashed line represents the Silhouette score and it is evident that the clustering quality declines as the number of clusters increases.

Figure 8: This image displays the results of the Principal Component Analysis (PCA) in two dimensions on the analyzed dataset, highlighting the two clusters (red and green) identified by our algorithm.

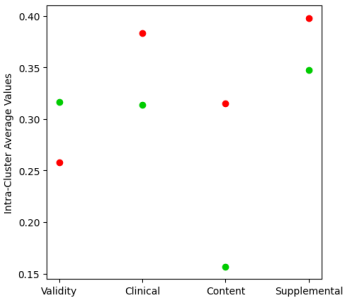

Figure 9: This image displays the four main group scales (Validity, Clinical, Content, and Supplemental) on the x-axis, and the y-axis presents the intra-cluster average values for each of these psychological scales for the two clusters (red and green) identified by our algorithm.

in Fig. 11, where the x-axis represents the average values of the Content scale and the y-axis the average values of the Clinical scale for each element in the dataset.

Finally, Fig. 10 provides a deeper analysis of the weights associated with the psychological scales for the first and second principal components of the PCA. From this plot, it is clear that for the elements in the green cluster, high values on scales related to the Content group correspond to highly positive weights, while low values correspond to negative weights. In contrast, the red cluster exhibits an inverted trend. For the second principal component, the red cluster elements are more evenly distributed across the dimension, while the green cluster elements generally show lower values across the scales. From these graphs, psychology experts can gain insights into the most relevant psychological scales within the dataset, thereby speeding up and simplifying the initial data analysis.

7. Conclusion

In this study, we proposed a novel approach for analyzing MMPI-2 profiles of prospective adoptive parents using evolutionary clustering techniques. By incorporating a genetic algorithm to autonomously select the most relevant psychometric scales, we aimed to streamline the clustering process and reduce reliance on manual selection by domain experts.

By employing a genetic algorithm to automatically select the most relevant psychological scales, combined with K-Means clustering based on minimum centroid distance and Silhouette analysis, we determined that two clusters were the optimal choice to describe the analyzed dataset.

These clusters displayed distinct psychological profiles, with notable differences particularly in the content and clinical scales, which may serve as valuable insights for

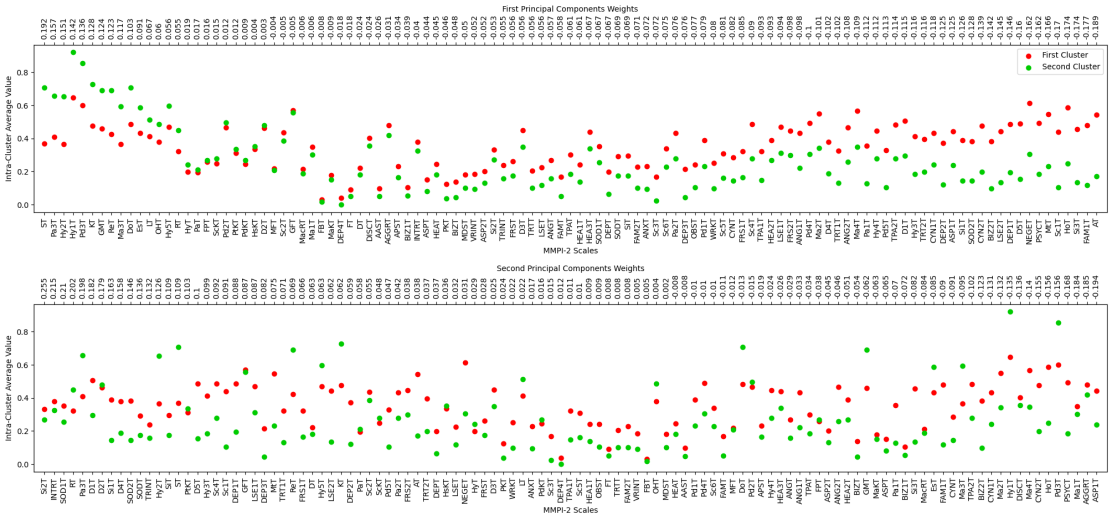

Figure 10: These two plots display, for the two clusters identified by our algorithm (red and green), the intra-cluster average value for each single psychological scale in the dataset on the y-axis, and the weights associated with the first principal component in the top plot and the second principal component in the bottom plot on the x-axis.

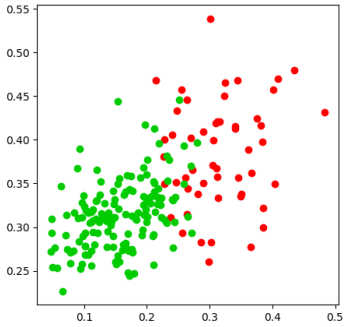

Figure 11: This image shows the average values of the content scale on the x-axis and the clinical scale on the y-axis for each element in the dataset. The division along the x-axis is clearly visible, while on the y-axis, elements in the green cluster tend to be more concentrated, whereas the red cluster elements are more evenly distributed across the clinical scale.

psychologists when assessing potential adopters.

The implications of our approach are twofold: first, it offers a data-driven methodology that enhances the initial interpretation of complex MMPI-2 profiles, assisting psychologists in identifying meaningful patterns without prior assumptions. Second, it underscores the potential of unsupervised learning techniques, such as genetic algorithms, in improving psychometric data analysis by automating feature selection and optimizing clustering quality.

Future work may involve expanding the dataset and further refining the genetic algorithm to handle larger and more diverse MMPI profiles. Additionally, exploring the integration of other clustering methods and incorporating newer versions of the MMPI test, such as MMPI-3, may provide further improvements and adaptability in diverse psychological evaluations.

References

[1] D. E. Johnson, Adoption and the effect on children's development, Early human development 68 (2002) 39–54.

[2] L. M. Beauchesne, As if born to: The social construction of a deficit identity position for adopted persons (1997).

[3] M. DeJong, J. Hodges, O. Malik, Children after adoption: Exploring their psychological needs, Clinical child psychology and psychiatry 21 (2016) 536–550.

[4] D. Brodzinsky, M. Gunnar, J. Palacios, Adoption and trauma: Risks, recovery, and the lived experience of adoption, Child abuse & neglect 130 (2022) 105309.

[5] S. Russo, S. I. Illari, R. Avanzato, C. Napoli, Reducing the psychological burden of isolated oncological patients by means of decision trees, in: CEUR Workshop Proceedings, volume 2768, 2020, p. 46 – 53.

[6] G. Lo Sciuto, S. Russo, C. Napoli, A cloud-based flexible solution for psychometric tests validation, administration and evaluation, in: CEUR Workshop Proceedings, volume 2468, 2019, p. 16 – 21.

[7] S. Russo, C. Napoli, A comprehensive solution for

psychological treatment and therapeutic path planning based on knowledge base and expertise sharing, in: CEUR Workshop Proceedings, volume 2472, 2019, p. 41 – 47.

[8] S. I. Illari, S. Russo, R. Avanzato, C. Napoli, A cloud-oriented architecture for the remote assessment and follow-up of hospitalized patients, in: CEUR Workshop Proceedings, volume 2694, 2020, p. 29 – 35.

[9] S. Hathaway, J. McKinley, A multiphasic personality schedule (minnesota): I. construction of the schedule, The Journal of Psychology 10 (1940) 249–254.

[10] J. N. Butcher, Minnesota Multiphasic Personality Inventory-2: Manual for administration, scoring, and interpretation, University of Minnesota Press, 2001.

[11] J. Butcher, C. Williams, J. Graham, A. Tellegen, Y. Ben-Porath, R. Archer, et al., Manual para la administración, puntaje e interpretación del inventario multifacético de la personalidad en adolescentes, mmpi-a [manual for administration, scoring, and interpretation of the minnesota multiphasic personality inventory for adolescents: Mmpi-a], 1992.

[12] Y. Ben-Porath, A. Tellegen, Minnesota Multiphasic Personality Inventory-3 (MMPI-3): Manual for administration, scoring, and interpretation, University of Minnesota Press, 2020.

[13] R. Xu, D. Wunsch, Survey of clustering algorithms, IEEE Transactions on neural networks 16 (2005) 645–678.

[14] M. Woźniak, D. Połap, R. K. Nowicki, C. Napoli, G. Pappalardo, E. Tramontana, Novel approach toward medical signals classifier, in: Proceedings of the International Joint Conference on Neural Networks, volume September 2015, 2015. doi:10.1109/IJCNN.2015.7280556.

[15] E. Iacobelli, V. Ponzi, S. Russo, C. Napoli, Eye-tracking system with low-end hardware: Development and evaluation, Information (Switzerland) 14 (2023). doi:10.3390/info14120644.

[16] F. Fiani, S. Russo, C. Napoli, An advanced solution based on machine learning for remote emdr therapy, Technologies 11 (2023). doi:10.3390/technologies11060172.

[17] I. E. Tibermacine, A. Tibermacine, W. Guettala, C. Napoli, S. Russo, Enhancing sentiment analysis on seed-iv dataset with vision transformers: A comparative study, in: ACM International Conference Proceeding Series, 2023, p. 238 – 246. doi:10.1145/3638985.3639024.

[18] C. Napoli, G. Pappalardo, E. Tramontana, Z. Marszalek, D. Polap, M. Wozniak, Simplified firefly algorithm for 2d image key-points search, in: IEEE SSCI 2014 - 2014 IEEE Symposium Series on Computational Intelligence - CIHLI 2014: 2014 IEEE Symposium on Computational Intelligence for Human-Like Intelligence, Proceedings, 2014. doi:10.1109/CIHLI.2014.7013395.

[19] M. Woźniak, D. Połap, M. Gabryel, R. K. Nowicki, C. Napoli, E. Tramontana, Can we process 2d images using artificial bee colony?, in: Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science), volume 9119, 2015, p. 660 – 671. doi:10.1007/978-3-319-19324-3_59.

[20] J. Shapiro, Genetic algorithms in machine learning, in: Advanced Course on Artificial Intelligence, Springer, 1999, pp. 146–168.

[21] M. Woźniak, D. Połap, C. Napoli, E. Tramontana, Graphic object feature extraction system based on cuckoo search algorithm, Expert Systems with Applications 66 (2016) 20 – 31. doi:10.1016/j.eswa.2016.08.068.

[22] K. P. Sinaga, M.-S. Yang, Unsupervised k-means clustering algorithm, IEEE access 8 (2020) 80716–80727.

[23] V. Ponzi, S. Russo, A. Wajda, R. Brociek, C. Napoli, Analysis pre and post covid-19 pandemic rorschach test data of using em algorithms and gmm models, in: CEUR Workshop Proceedings, volume 3360, 2022, p. 55 – 63.

[24] N. Brandizzi, S. Russo, G. Galati, C. Napoli, Addressing vehicle sharing through behavioral analysis: A solution to user clustering using recency-frequency-monetary and vehicle relocation based on neighborhood splits, Information (Switzerland) 13 (2022). doi:10.3390/info13110511.

[25] M. Ahmed, R. Seraj, S. M. S. Islam, The k-means algorithm: A comprehensive survey and performance evaluation, Electronics 9 (2020) 1295.

[26] N. Andriyanov, A. Tashlinsky, V. Dementiev, Detailed clustering based on gaussian mixture models, in: Intelligent Systems and Applications: Proceedings of the 2020 Intelligent Systems Conference (IntelliSys) Volume 2, Springer, 2021, pp. 437–448.

[27] N. Brandizzi, S. Russo, R. Brociek, A. Wajda, First studies to apply the theory of mind theory to green and smart mobility by using gaussian area clustering, in: CEUR Workshop Proceedings, volume 3118, 2021, p. 71 – 76.

[28] U. Von Luxburg, A tutorial on spectral clustering, Statistics and computing 17 (2007) 395–416.

[29] V. Meuti, I. Marini, A. Grillo, M. Lauriola, C. Leone, N. Giacchetti, F. Aceti, et al., Mmpi-2: cluster analysis of personality profiles in perinatal depression—preliminary evidence, The Scientific World Journal 2014 (2014).

[30] J. MacQueen, et al., Some methods for classification and analysis of multivariate observations, in:

Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, volume 1, Oakland, CA, USA, 1967, pp. 281–297.
[31] J. L. RILEY, M. E. Robinson, M. E. Geisser, V. T. Wittmer, Multivariate cluster analysis of the mmpi-2 in chronic low-back pain patients, 1993.
[32] J. D. Elhai, B. C. Frueh, Subtypes of clinical presentations in malingerers of posttraumatic stress disorder: An mmpi-2 cluster analysis, Assessment 8 (2001) 75–84.
[33] L. E. Aguerrevere, M. R. Calamia, K. W. Greve, K. J. Bianchini, K. L. Curtis, V. Ramirez, Clusters of financially incentivized chronic pain patients using the minnesota multiphasic personality inventory-2 restructured form (mmpi-2-rf)., Psychological Assessment 30 (2018) 634.
[34] K. De Jong, Learning with genetic algorithms: An overview, Machine learning 3 (1988) 121–138.
[35] K. De Jong, Adaptive system design: a genetic approach, IEEE Transactions on Systems, Man, and Cybernetics 10 (1980) 566–574.
[36] J. J. Grefenstette, Optimization of control parameters for genetic algorithms, IEEE Transactions on systems, man, and cybernetics 16 (1986) 122–128.
[37] K. A. De Jong, An analysis of the behavior of a class of genetic adaptive systems., University of Michigan, 1975.
[38] H. AL-Tabtabai, A. P. Alex, Using genetic algorithms to solve optimization problems in construction, Engineering Construction and Architectural Management 6 (1999) 121–132.
[39] K. Deb, Genetic algorithm in search and optimization: the technique and applications, in: Proceedings of International Workshop on Soft Computing and Intelligent Systems,(ISI, Calcutta, India), Proceedings of International Workshop on Soft Computing and Intelligent ..., 1998, pp. 58–87.
[40] A. Homaifar, C. X. Qi, S. H. Lai, Constrained optimization via genetic algorithms, Simulation 62 (1994) 242–253.
[41] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the em algorithm, Journal of the royal statistical society: series B (methodological) 39 (1977) 1–22.
[42] H. Abdi, L. J. Williams, Principal component analysis, Wiley interdisciplinary reviews: computational statistics 2 (2010) 433–459.